SYNTHESIS OF UNNATURAL TRYPTOPHAN ANALOGS TO INVESTIGATE THE MOLECULAR RECOGNITION OF METHYLATED LYSINE BY THE HP1 CHROMODOMAIN

M. Kaitlyn Tsai

Honors Thesis

May 2016

ABSTRACT

DNA is wrapped around a histone protein bundle, and these histone proteins can undergo post-translational modifications which then participate in signaling pathways to regulate important cellular functions. One such histone modification is methylation, where lysine can be mono, di, or tri-methylated. Reader proteins bind to methyl lysine and serve as docking sites for additional proteins that alter gene expression or chromatin structure. Dysregulation of histone methylation has been associated with certain types of cancer; therefore, it is necessary to understand the driving forces for binding of methylated lysine by these reader proteins to aid the development of reader protein probes or inhibitors with the desired specificity. In this project, the cation-π interaction between trimethylated lysine 9 (Kme_3) of H3 and the *Drosophila* reader protein, HP1 chromodomain, is investigated. The binding site of HP1 is comprised of three aromatic amino acids, two tyrosine residues and a tryptophan. To better understand the contribution of tryptophan in this cation- π binding event, an unnatural tryptophan analog is synthesized with 1 fluorine atom, which is highly electron-withdrawing, attached to the outer aromatic rung. Once synthesized, these unnatural analogs will then be inserted into the HP1 binding site. Comparative binding studies between the wild type and mutant protein with the Kme_3 peptide will provide a better understanding of the cation- π mechanism. If Kme_3 binds less tightly to the mutated binding pockets, it will indicate the importance of the tryptophan- π contribution to binding.

INTRODUCTION

Genetics refers to the study of how organisms change in response to alterations to their genetic code. However, epigenetics, where the prefix “epi” means *above* or *over,* refers to the study of how organisms change in response to alterations to gene expression and does not

involve modification to the DNA sequence. Although advancements in genetic research and technologies have sky-rocketed over the past few decades, epigenetics remains an area that is poorly understood despite its importance to biological processes. For example, epigenetics is central to cellular differentiation. Every cell in our body, with the exception of a select few, has an identical genetic code; however, we maintain a variety of cell types in our bodies. Epigenetic mechanisms dictate which genes are silenced or expressed to give cells a particular function. Improper gene expression is also implicated in certain cancers and developmental disorders, making the investigation into epigenetic malfunction relevant to human health.

An important type of epigenetic mechanism is histone modification. DNA is wrapped around a bundle of 8 histone proteins (2 copies each of H2A, H2B, H3 and H4).[1] Complexes of DNA and protein, referred to as chromatin, can then undergo further compaction into chromosomes to fit inside cell nuclei. Each histone protein maintains unstructured regions on its N-terminus. Proteins called "writers" subject N-terminal residues to various post-translational modifications, including methylation, acetylation, ubiquitination. "Eraser" proteins can remove these modifications, and proteins called "readers" can recognize these modifications and participate in signaling pathways to influence cellular function.

One type of post-translational modification involves the methylation of histones H3 and H4, where the amino acid lysine can be mono, di, or tri-methylated. Reader proteins bind sequence-selectively to methyl lysine to serve as docking sites for additional proteins involved in gene expression or chromatin structure. Dysregulation of lysine methylation in histones has been associated with certain types of cancers and is a new potential drug target.[2] For example, it is thought that lacking trimethylated lysine 9 on histone 3 in mice results in genome instability and tumor progression.

My project is primarily concerned with trimethylated lysine (Kme_3). There are several types of reader proteins that recognize Kme_3. Most of these reader proteins have similar aromatic cage binding pockets but differ in the number, orientation, and type of participating aromatic residues. It is important to understand what interactions are most important for Kme_3 binding to aid the design of reader protein probes or inhibitors with the desired specificity. Because reader proteins are important for epigenetic pathways, probes could help uncover their role in larger epigenetic contexts, and inhibitors could be utilized as epigenetic therapy.[3]

For these reader proteins, it is believed that the driving force for binding and selectivity of the aromatic binding pocket to the positively-charged Kme_3 is a cation-π interaction.[1] Cation-π interactions are non-covalent interactions that are prevalent in biological molecular recognition systems.[4] This type of interaction involves a cation interacting with an electron-rich face of an aromatic molecule and is worth 1-5 kcal/mol in water, which is comparable with hydrogen bonding and electrostatic interactions.

My project is focused on the *Drosophila* reader protein, heterochromatin protein 1 (HP1) chromodomain, which selectively recognizes Kme_3 on histone 3 (Figure 1). Like other reader proteins that also recognize Kme_3, its binding pocket is an aromatic cage, consisting of 2 tyrosine residues and 1 tryptophan. In a previous study, the Waters group demonstrated the importance the cation-π interaction for the binding of HP1 to Kme_3.[1] Methylated lysine was substituted for an unnatural amino acid in which the nitrogen of lysine was replaced with carbon, providing similar shape and van der Waals interactions but neutral charge. Comparative binding studies found that the HP1's affinity for the

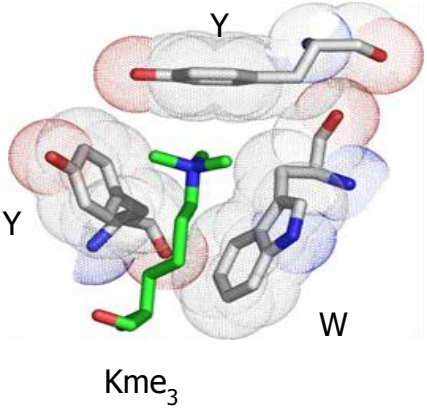

Figure 1. Binding pocket of the *Drosophila* HP1 chromodomain in complex with trimethylated lysine (Kme3).

neutral compound was reduced by a factor of 30. This result supports that a cation-π interaction, rather than the hydrophobic effect, is the driver for favorable binding.

Along similar lines, my project aims to increase understanding the π component of this cation- π interaction between the HP1 and Kme_3 by determining if all three aromatic residues contribute equally to binding. My project sought to determine the importance of tryptophan for this interaction. To accomplish this task, I worked to synthesize 5-fluoro-tryptophan, which is natural tryptophan with a fluorine atom attached to the outer ring (Figure 2). Once this unnatural analog is successfully synthesized, it, along with other tryptophan analogs with higher order fluorine substitutions, will be incorporated into the binding site of HP1. Fluorine was chosen due to its small size, to limit steric effects, and its electron-withdrawing properties. The addition of fluorine to tryptophan decreases the electron density on the aromatic face. Therefore if tryptophan is an important π contributor in this cation- π interaction, mutant HP1 binding pockets containing fluorinated tryptophan should have a decreased affinity for Kme_3.

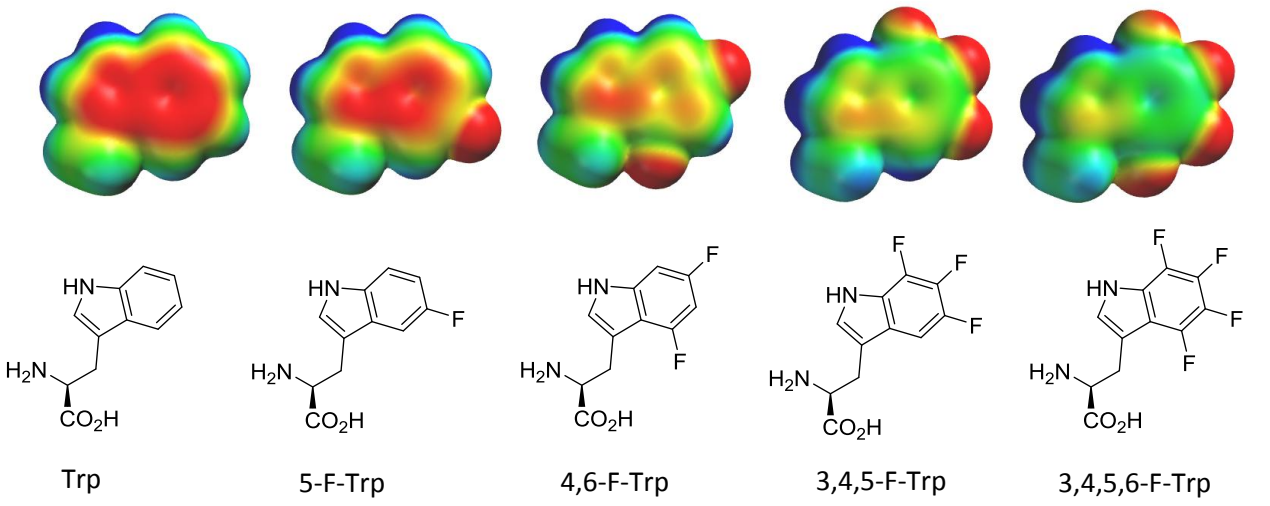

Figure 2. Electrostatic potential maps of natural tryptophan (left), 5-fluoro-tryptophan (synthetic target) and tryptophan analogs with 2, 3, and 4 fluorine additions to the aromatic ring. Red indicates high electron density, green is neutral, and blue is low electron density.

METHODS

Synthesis of 5-fluoro tryptophan was achieved through previously established methods (Figure 3).[5] Using Vilsmeier-Haack conditions, the fluorinated indole is first formylated. Then, using sodium hydride for deprotonation, the amine is protected with a tosyl group. The aldehyde is then reduced to the alcohol using sodium borohydride, which is then brominated using N-bromosuccinimide and triphenylphosphine. After I had obtained the bromide intermediate, I utilized two different chiral auxiliaries to attempt asymmetric synthesis of the α-amino acid, a proline-derived Ni(II) complex and the Schöllkopf reagent. For both chiral auxiliaries, the alkylation product is then subjected to similar acid or acid/base hydrolysis to give the desired α-amino acid.

Using the Ni(II) complex for asymmetric synthesis was attractive since it could be prepared form inexpensive starting materials, alkylation could be achieved using mild conditions with high stereoselectivity, and part of the Ni(II) complex could be regenerated after hydrolysis.[6] However, preparation of the Ni(II) complex proved difficult. The Schöllkopf reagent, unlike the Ni(II) complex, is commercially available and also achieves high stereoselectivity. However, the Schöllkopf reagent is expensive, alkylation must be conducted under harsher conditions, and the chiral auxiliary cannot be regenerated.

I synthesized the proline-derived Ni(II) complex using previously established methods (Figure 4).[7] Sodium methoxide and benzyl chloride are used to achieve the mono-benzylation of proline. Then, using 1-methyl imidazole as a catalyst/base and methanesulfonyl chloride as an activator, an amide bond is formed. Lastly, heating this compound in the presence of glycine, $Ni(NO_2)_2$, methanol, and sodium methoxide gives the desired Ni(II) complex.

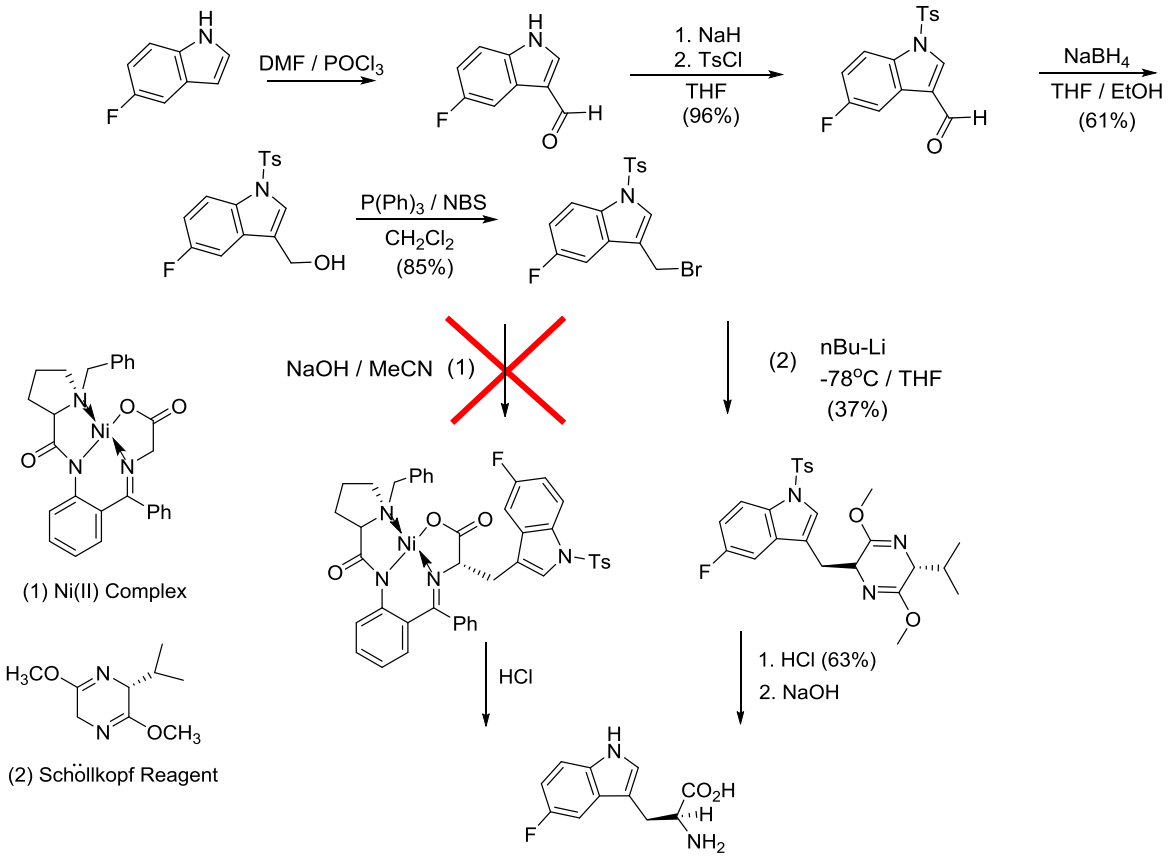

Figure 3. Synthetic scheme for 5-fluoro-tryoptophan, utilizing two different chiral auxiliaries to achieve asymmetric synthesis of the α-amino acid.[5,7]

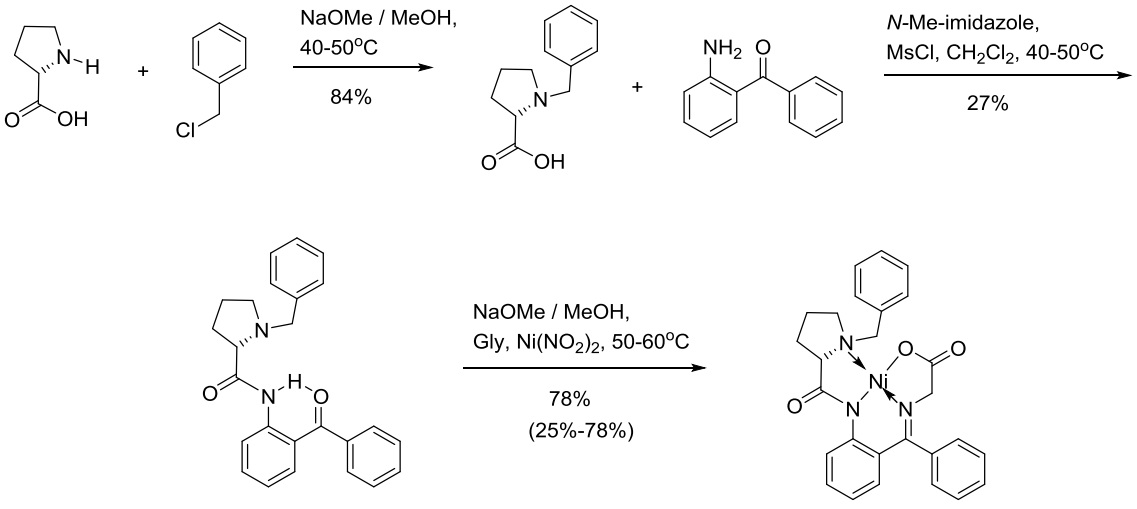

Figure 4. Synthesis of the proline-derived Ni(II) complex for the asymmetric synthesis of the α-amino acid, 5-fluoro-tryptophan.[7]

RESULTS

For the synthesis of 5-fluoro tryptophan (Figure 3), I obtained acceptable yields for the first four reactions to obtain the bromide intermediate, a dark brown oil (Supp. Data, Figure 5). I obtained 96% yield over the first two reactions, 61% for the reduction step, and 85% for the bromination step.

For the preparation of the proline-derived Ni(II) complex, I was able to obtain acceptable yields for the first benzylation step (84%). For the second step, I only achieved a 27% yield. Although I achieved a 78% yield for the last step, I was only able to obtain high yields once by altering the purification protocol. The literature indicated that purification should be conducted by silica-gel chromatography; however, I found that higher yields and purity could be accomplished by recrystallization to obtain red crystals (Supp. Data, Figure 6).

Unfortunately, utilization of the proline-derived Ni(II) complex for alkylation proved unsuccessful (Supp. Data, Figure 7). To evaluate if steric hindrance was preventing proper alkylation, I conducted a test alkylation using 3-fluoro-4-methoxybenzyl bromide, which is smaller, cheap, and commercially available (Figure 8). Furthermore, this compound maintains similar features to the bromide intermediates I had synthesized, including the presence of a fluorine and aromatic component. However, alkylation using this reagent was also unsuccessful (Supp. Data, Figure 9), indicating that perhaps other factors such as electronic effects were preventing proper alkylation.

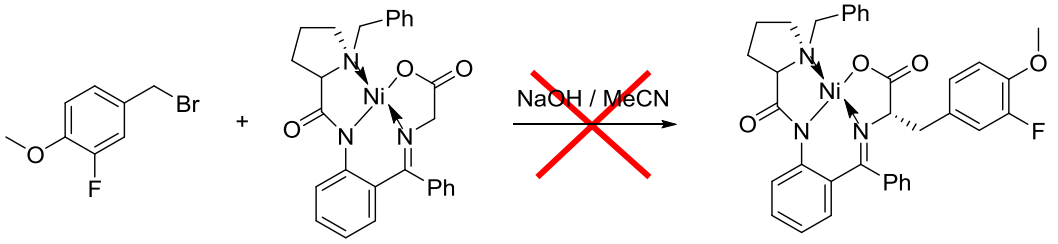

Figure 8. Unsuccessful test alkylation using 3-fluoro-4-methoxybenzyl bromide and the proline-derived Ni(II) complex.

Since the Ni(II) complex proved difficult to synthesize and alkylation using the synthesized bromide intermediates and a less-bulky analog were both unsuccessful, I attempted to achieve alkylation using the Schöllkopf reagent instead. I first conducted a test alkylation using the 3-fluoro-4-methoxybenzyl bromide (Figure 10). I was able to obtain acceptable yields and pure product (Supp. Data, Figure 11).

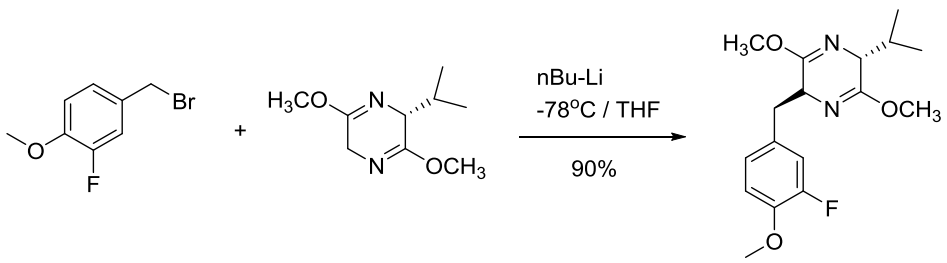

Figure 10. Test alkylation using 3-fluoro-4-methoxybenzyl bromide and the Schöllkopf Reagent.

Using the synthesized bromide intermediates, I achieved successful alkylation (37% yield) using the Schöllkopf reagent (Supp. Data, Figure 12). I also successfully performed acid hydrolysis to yield the ester product with a 63% yield. However, base hydrolysis with sodium hydroxide was unsuccessful, resulting in little or no desired α-amino acid product, 5-fluoro tryptophan.

DISCUSSION

To achieve insertion of unnatural tryptophan analogs into the HP1 binding site, large quantities of these amino acids must be synthesized. Therefore, reaction yields must be maximized. The two steps that tended to give the lowest yields were alkylation using the Schöllkopf reagent and base hydrolysis of the ester. For the alkylation step, different solvent systems could be tested for optimal conditions. For the base hydrolysis step, it may be beneficial to use a stronger base, such as lithium hydroxide to promote complete hydrolysis. I also observed that reaction yields in the summer were consistently less than yields during the winter months, which suggests that humidity may play a role. Because many of these reaction steps are water-sensitive, extra precautions may need to be taken in the summer months, such as flame-drying glassware and using new solvent bottles.

Once these steps have been optimized and scaled-up, 5-fluoro-tryptohan and other unnatural tryptophan analogs with higher-order fluorine substitutions will be incorporated into the HP1 binding pocket. Then, by comparing the affinity of mutant and wild-type HP1 for the positively-charged Kme_3 ligand, the importance of tryptophan in this cation$-\pi$ interaction can be evaluated. In the future, it would be beneficial to study the other aromatic tyrosine positions in the HP1 binding site by similar methods to gain a more complete understanding of the driving forces for binding and selectivity of the HP1 chromodomain for Kme_3.

For reader proteins that recognize methylated lysine, such as HP1, it is important to know which interactions are the major drivers for binding to aid the development of selective inhibitors or probes. These inhibitors and probes could then be utilized in future studies to help understand how histone methylation and their reader protein counterparts participate in the larger context of epigenetic regulation.

ACKNOWLEDGEMENTS

I would like to thank Dr. Marcey Waters for her continual support and advisement during the course of my undergraduate studies. She inspired me challenge myself and excel beyond what I believed was possible. Thank you, Amber Koenig, for investing countless lab hours to teach me the art of organic synthesis. I would also like to acknowledge Nate John, for always giving me both advice and a good laugh. I would also like to thank the entire Waters group, who were willing to answer my many questions and created a wonderful lab environment to work in. Lastly, I would like to thank the Barry M. Goldwater Scholarship and Frieze Summer Undergraduate Research Fellowship for providing funding to help complete this research.

REFERENCES

1. Huges, R.; Waters, M.;*PNAS* **2007,** *12,* 2786 – 2790.
2. Peters A.H.; *Cell.* **2001,** *107(3),* 323-37.
3. Seet. B.; *Nature Review Molecular Cell Biology.* **2006,** *7,* 473-483.
4. Dougherty, D.;. *A. Acc. Chem. Res.* **2013**, *46,* 885–893.
5. Konas, D; et. al; *Synth. Comm.* **2012,** *42,* 144-152.
6. Sorochinsky, A; et. al; *Amino Acids* **2013,** *45,* 691-718.
7. Soloshonok, V.; et. al; *J. Org. Chem.* **2003,** *68,* 7104- 7107.

SUPPLEMENTAL DATA

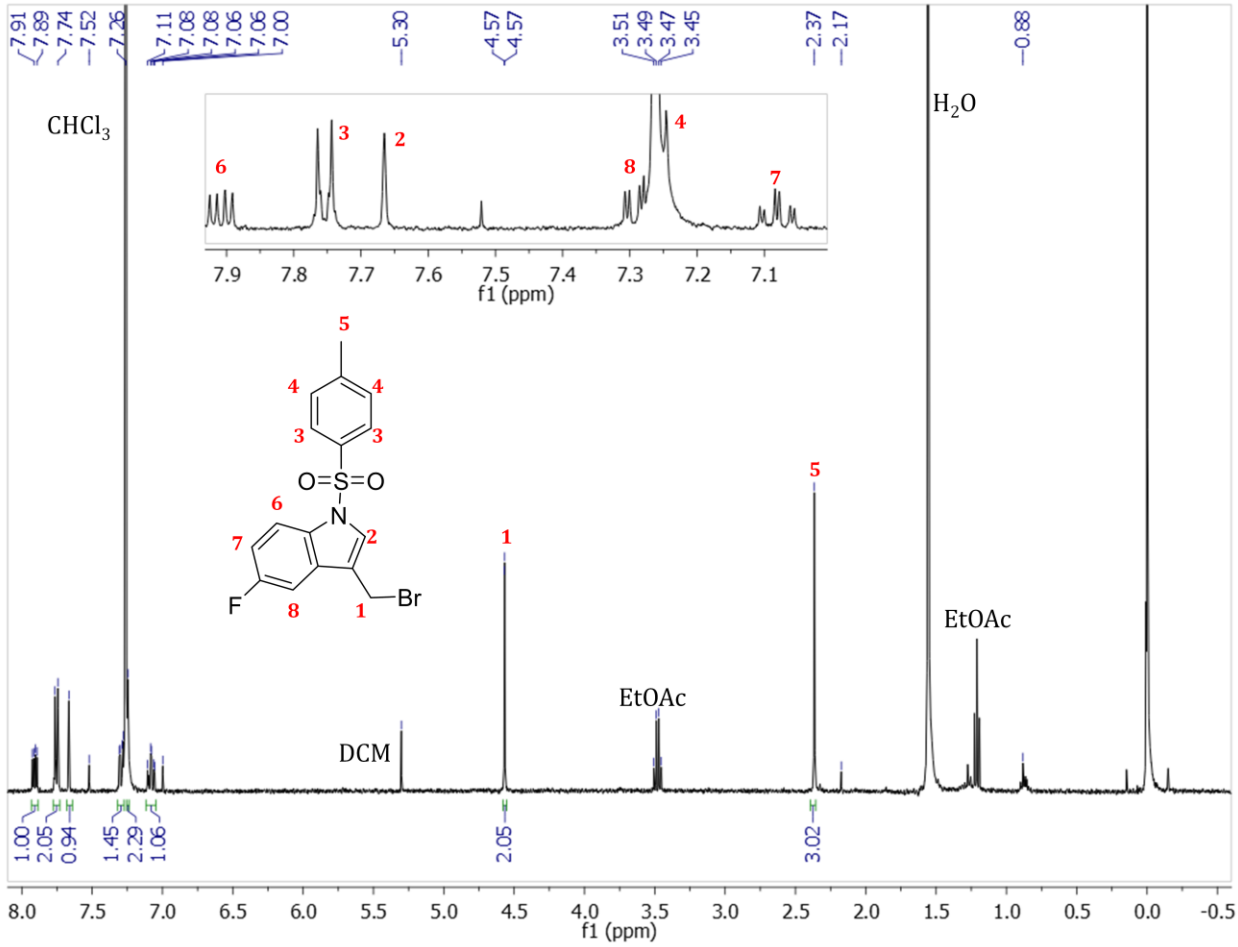

Figure 5. Annotated H^1 NMR spectrum of bromide intermediate (400MHz).

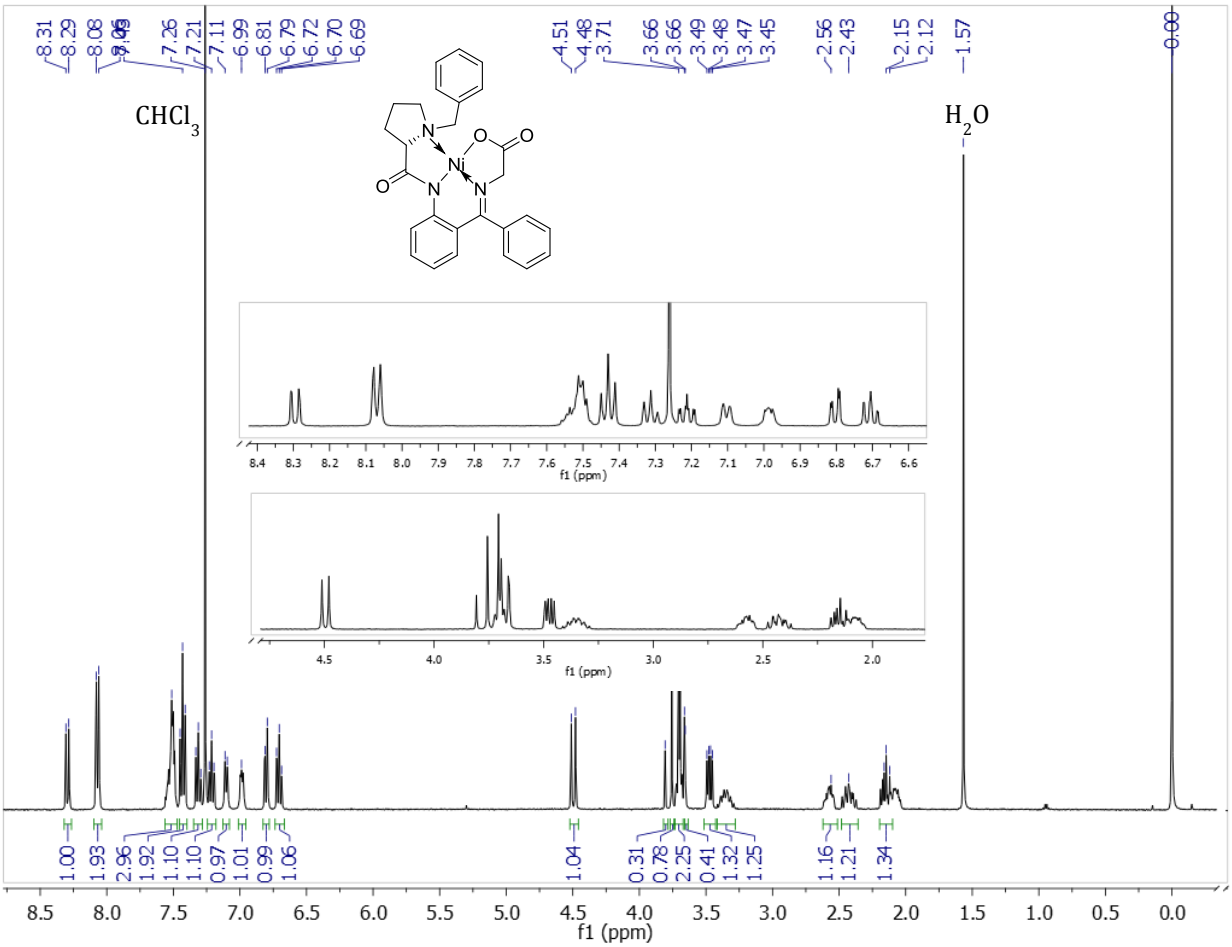

Figure 6. H^1 NMR spectrum of the proline-derived Ni(II) complex (400 MHz). The spectrum matches well with literature.[6]

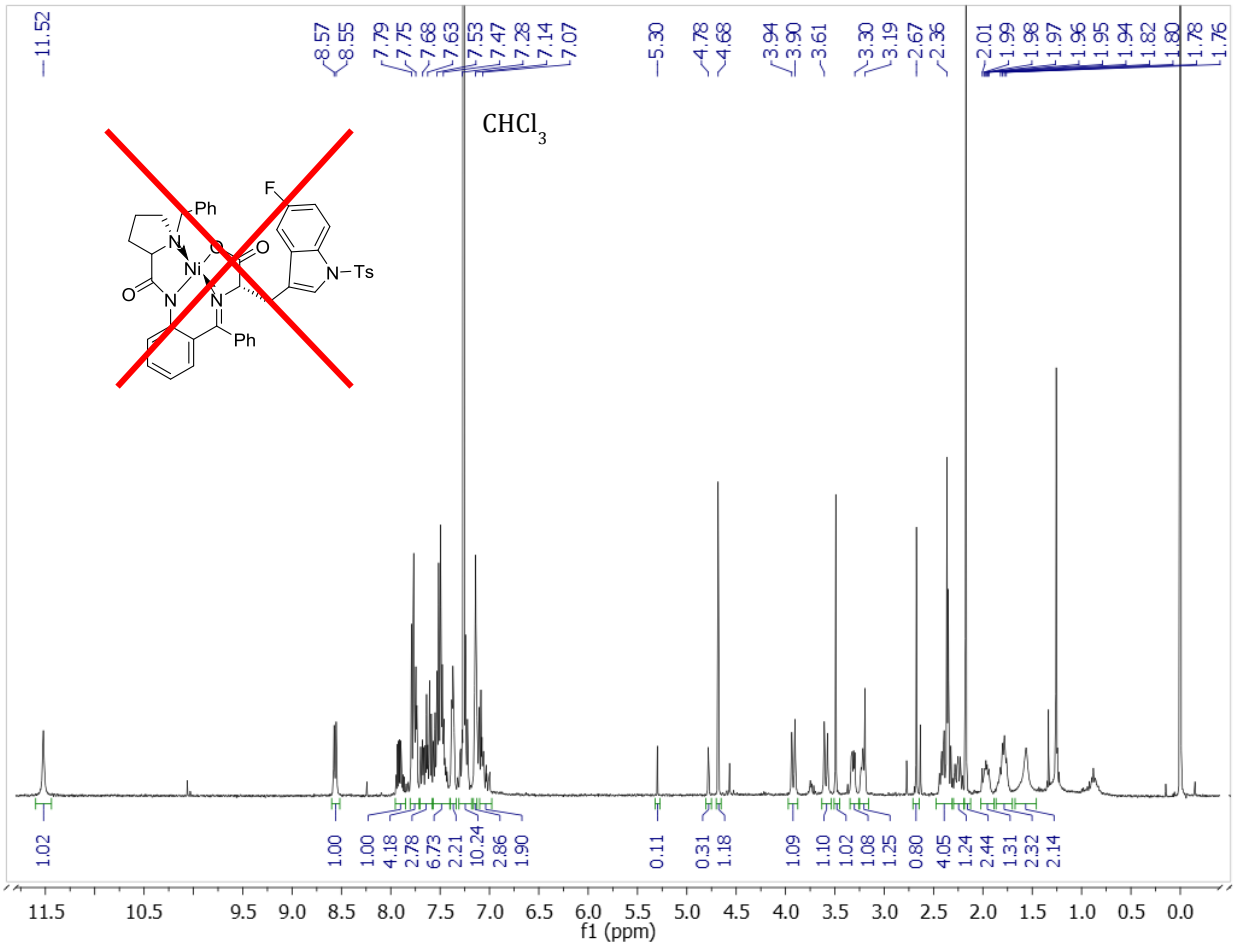

Figure 7. H^1 NMR spectrum (400 MHz) representing unsuccessful alkylation using the proline-derived Ni(II) complex to achieve asymmetric synthesis of the α-amino acid.

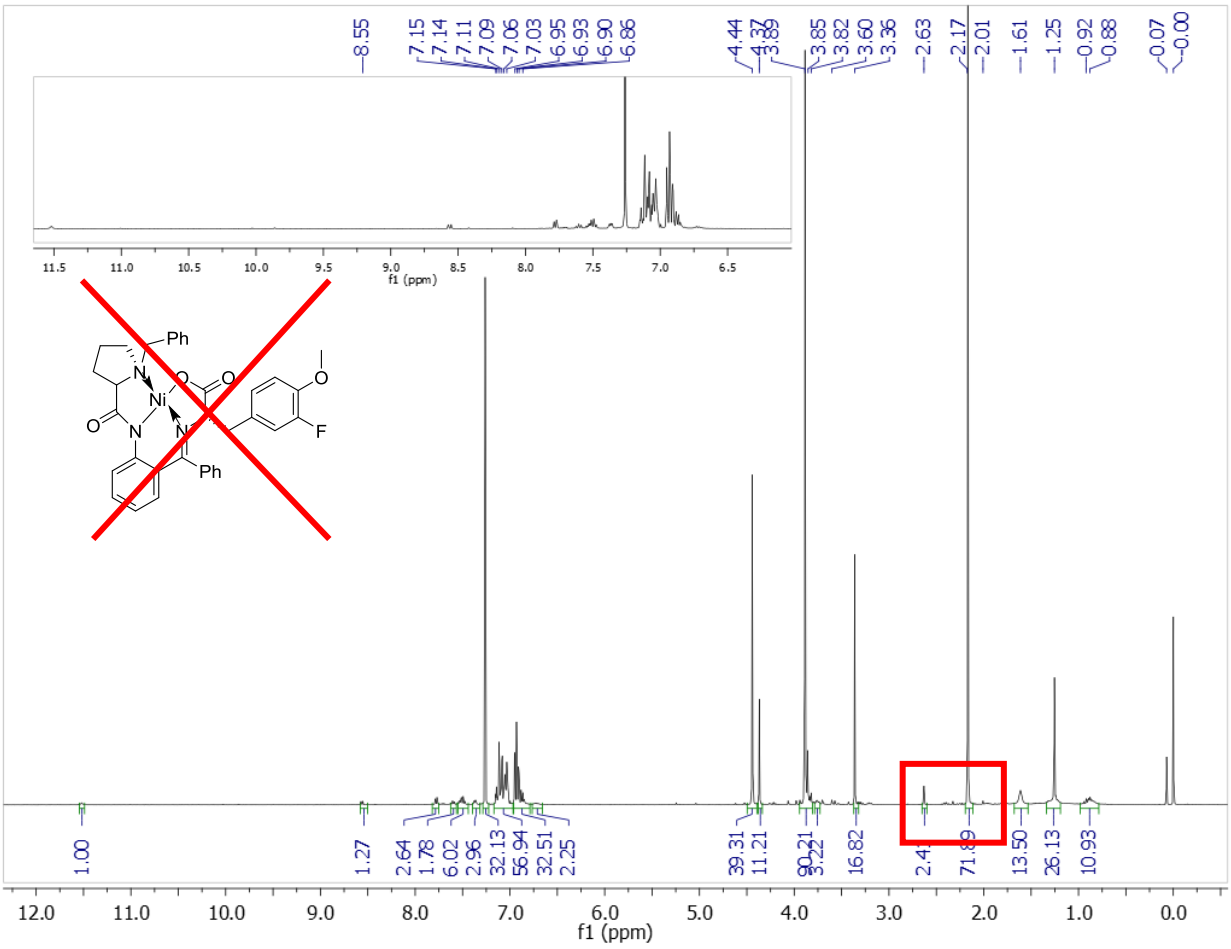

Figure 9. H^1 NMR spectrum (400 MHz) representing unsuccessful test alkylation using 3-fluoro-4-methoxybenzyl bromide and proline-derived Ni(II) complex. The red box highlights the absence of signal corresponding to proline protons.

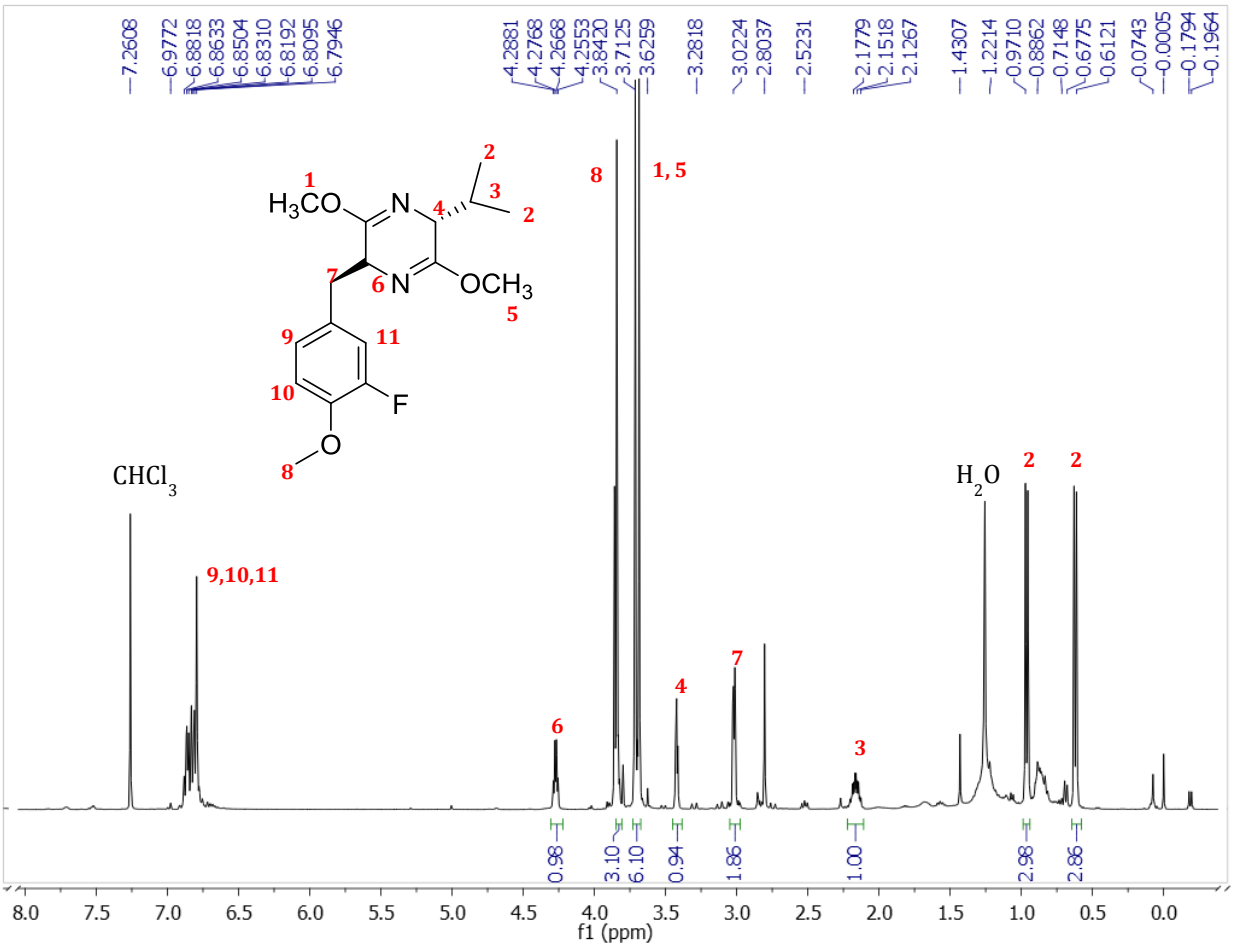

Figure 11. Annotated H^1 NMR spectra (400 MHz) of alkylation product of 3-fluoro-4-methoxybenzyl bromide and the Schöllkopf reagent.

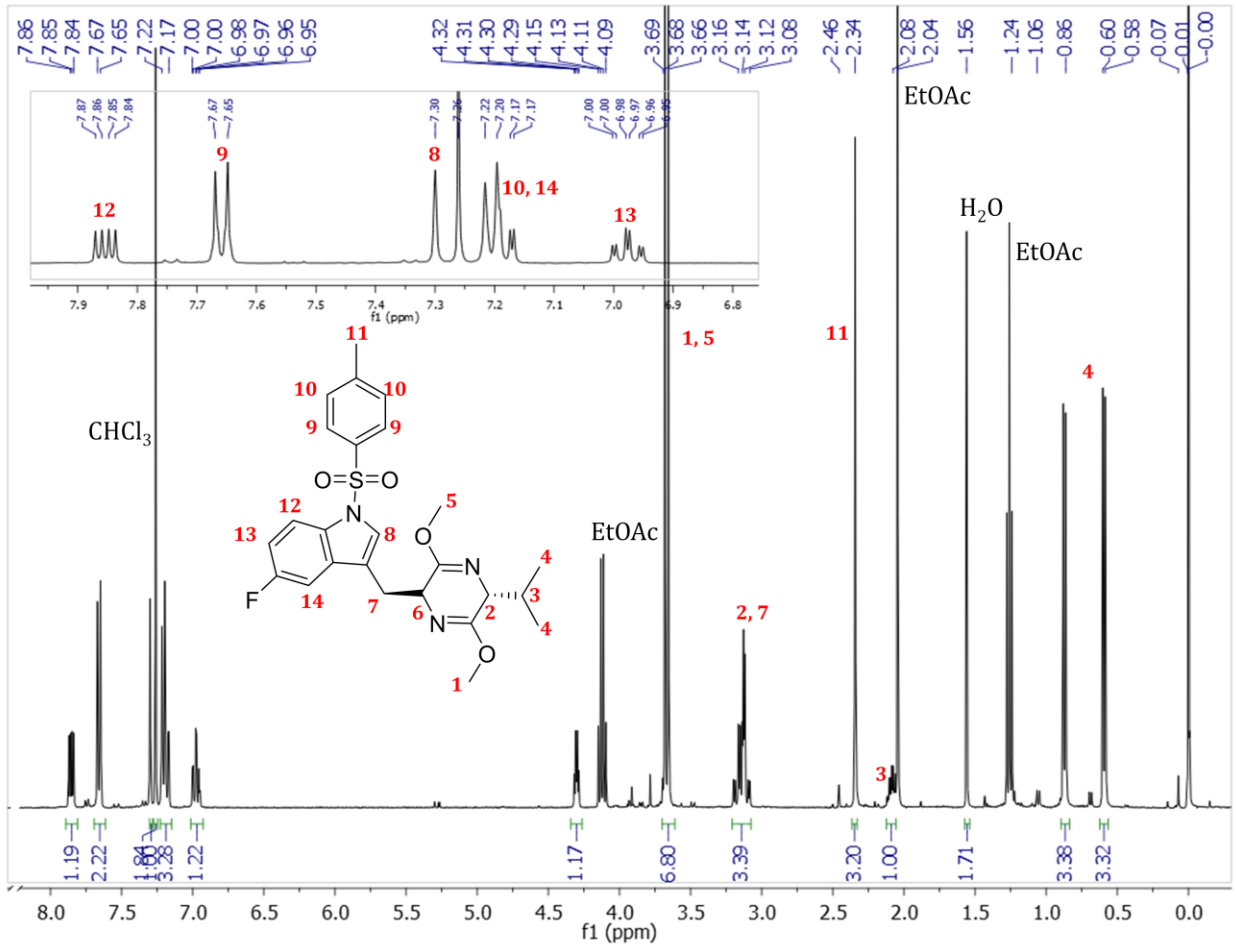

Figure 12. Annotated H^1 NMR spectra (400 MHz) of alkylation product of the fluorinated tryptophan bromide and Schöllkopf reagent.